

## Using Explainable AI to understand frailty indicators

Attayeb Mohsen<sup>a,\*</sup>, Masaki Yamamoto<sup>a</sup>, Agustin Martin-Morales<sup>a</sup>, Daiki Watanabe<sup>a,b</sup>, Nobuo Nishi<sup>a,c</sup>, Takashi Nakagata<sup>a</sup>, Tsukasa Yoshida<sup>a</sup>, Motohiko Miyachi<sup>a,b,\*</sup>, Kenji Mizuguchi<sup>a,d</sup>, Michihiro Araki<sup>a,e,f,g,\*</sup>

<sup>a</sup>National Institutes of Biomedical Innovation, Health and Nutrition, Ibaraki, Osaka, Japan

<sup>b</sup>Faculty of Sport Sciences Waseda University 2-579-15 Mikajima Tokorozawa-city Saitama 359-1192 Japan.

<sup>c</sup>Graduate School of Public Health St. Luke's International University Chuo-ku 104-0045 Tokyo Japan.

<sup>d</sup>Institute for Protein Research Osaka University Suita Osaka Japan.

<sup>e</sup>Graduate School of Medicine Kyoto University Kyoto Japan.

<sup>f</sup>Graduate School of Science Technology and Innovation Kobe University 1-1 Rokkodai Nada-ku Kobe Hyogo 657-8501 Japan.

<sup>g</sup>National Cerebral and Cardiovascular Center 6-1 Kishibe-Shinmachi Suita Osaka 564-8565 Japan.

---

### Abstract

**Introduction:** The prevalence of frailty is on the rise with the aging population and increasing life expectancy, which often is accompanied by comorbidities. Frailty can be effectively detected using Frailty index such as KCL index. Early detection of frailty allows applying measures that reduce the conversion rate to frail, and improve the quality of life in the frail people. Therefore, to facilitate the screening of frailty status at the primary care level, we suggest to produce a shorter version of the KCL questionnaire.

**Aim:** To understand the importance of KCL components in the decision making process for frailty and use machine learning approach to shorten the Questionnaire while maintaining reasonable accuracy, making it easier to screen for frailty in primary care.

**Methods:** We developed an automated framework of three steps: Feature importance determination using Shap values, testing models with Cross-validation with increased addition of selected features. Moreover, we validated the reliability of KCL to detect

---

\*Corresponding authors

Email addresses: [attayeb@nibiohn.go.jp](mailto:attayeb@nibiohn.go.jp) (Attayeb Mohsen), [miyachim@waseda.jp](mailto:miyachim@waseda.jp) (Motohiko Miyachi), [araki@nibiohn.go.jp](mailto:araki@nibiohn.go.jp) (Michihiro Araki)

frailty by comparing the results of KCL criteria with the unsupervised clustering of the data. **Results:** Our approach allowed us to identify the most important questions in the KCL questionnaire and demonstrate its performance using a short version with only four questions (4) Do you visit homes of friends?, (6) Are you able to go upstairs without using handrails or the wall for support? (10) Do you feel anxious about falling when you walk?, and (25) (In the past two weeks) Have you felt exhausted for no apparent reason?). We also showed that the data clustering corresponds well with the results of KCL criteria. **Discussion and Conclusion:** While it is difficult to predict pre-frail status using shorter KCL questionnaire, it was shown to be fairly accurate in predicting frail status using only four questions.

---

## 1. Introduction

A frail person can be described as a person who is barely doing well, which implies that the person will break and his health will collapse as soon as they are exposed to even minor stress or an accident, such as falling or mild infection. Therefore, frailty can be defined as the antonym of fitness/fit[1, 2]. The course of frailty has been described as “a decline in functioning across multiple physiological systems, accompanied by an increased vulnerability to stressors”[3]. Studies have also shown that frail individuals are more likely to experience falls, hospitalization, and death[4, 5], which can greatly impact their quality of life and put a strain on the healthcare system and economy[6].

It is possible for frail people to transit to robust or the other way around. That was corroborated by Japan’s geriatric society when they changed their understanding of frailty from the natural decline of a person’s fitness, health, and psychological well-being, to a preventable condition that needs attention and cure in 2014[7].

Diagnosis of frailty should start at the primary care level, where the patient is typically treated comprehensively. The early assessment of frailty makes early detection of frailty status possible, which subsequently help prevent conversion of robust adults to frail, as well as facilitate the returning from frail to robust[8] by applying appropriate preventive and treatment measures. Over five years of follow-up, adults with low body mass index and more physical activity are more likely to remain robust than the others, suggesting that preventing conversion to frail by modifying their exercise behavior and dietary habits is possible [9].

Frailty assessment is built on two main models: (1) the phenotype model and its evaluation method is Cardiovascular Health Study (CHS) criteria created by Fried et al. [4], and (2) the accumulation deficit model and its representative evaluation method, the Frailty Index (FI), advocated by Rockwood et al. [10].

Fried et al. built their Cardiovascular Health Study (CHS) Criteria on the presence of three out of these five components: Shrinkage (weight loss), Weakness, Poor en-

durance, slowness, and low physical activity level [4].

The Asia-Pacific Clinical Practice Guidelines for Management of Frailty evaluated multiple frailty measurements and classified them based on their use in clinical practice. The Frailty Index (FI) and Kihon Checklist were both classified as assessment tools that not only assess physical domains, but also comprehensively include mental or psychological domains. [2]. A good frailty index can be created by counting the number of symptoms or signs a subject has, the frailty status of that subject correlates with that number. Moreover, the frailty index is consistent even if the symptoms or signs are not the same[11].

The Kihon Checklist (KCL) system [12, 13], that consists of 25-items, was developed and published by Japanese Ministry of Health, Labour and Welfare (MHLW) in 2006[14, 15]. Although KCL was developed initially to predict if individual requires long-term care, it has been used extensively as an evaluation of frailty due to its similarity to FI and its comparable results with CHS[16].

Recent research extensively evaluated the KCL system, and found it useful in predicting the need for long-term care, and the mortality risk. It also shows that KCL has higher accuracy among other frailty tools[17]. It revealed that the cutoff value of 7 points or more of 25 to define frailty is appropriate[16, 18, 19]. Currently, KCL is globally used and translated into various languages[12, 20–24].

Previous studies have employed machine learning to predict frailty using data of specific questionnaire with other measurements such as handgrip strength and walking speed[25]. They suggest that their models is particularly important in detecting the frailty in early stages, to facilitate prevention of further deterioration. Hassler et al. built predictive models for frailty and applied feature selection using the Boruta algorithm. They also examined the model performance by applying a 10-fold cross-validation approach. The dataset analyzed here has more than 200 variables. They fused the class pre-frail and frail into one class. Furthermore, they built a binary classification

model. Because the amount of missed data was significant in some variables, they needed to take many steps to preprocess the data, including imputing the missed values [26]. Aponte-Hao et al. examined the usage of machine learning to predict frailty in Canadian primary care clinics. They used a hold-out test set and examined eight supervised machine-learning algorithms. Their data included other features than our paper, such as diagnosis, billing codes, and medicine prescribed. Aponte-Hao et al. used a simple feature selection approach by removing the variables with less variation. Moreover, they removed the data with a high prevalence of missing values. The most important features were age, sex, and previous medical diagnosis. The sensitivity and specificity of the produced models were topped at 80% and 77%, respectively. [27]. The important point of these models is that they use regularly collected data in primary clinical care visits to predict frailty status. Kuo et al. built a machine-learning model to predict social frailty. Interestingly, they showed that one of the crucial predictors is religious participation (number 3) following health literacy and comorbidity. Religious participation item in here is similar to the KCL question about visiting a friend's house. They both reflect social and psychological well-being as well as musculoskeletal health [28]. Tseng et al. studied the prediction of cognitive frailty among the old Taiwanese cohort. Using logistic regression, they identified six factors associated with frailty, most related to age, gender and body weight, memory deficit, and medical history of diabetes mellitus [29].

More efforts are required to assess frailty and pull attention towards prevention and treatment[7]. New, more accessible approaches are required to diagnose frailty at the primary care level. Previously, Searle et al. suggested that using a high number of items in the frailty index leads to a more accurate diagnosis. However, this approach is built on the assumption that all symptoms are equally important on the assessment of frailty, while in reality their impact is sometimes different. In this study we will try a different approach, which is to use explainable AI to understand the importance of the frailty

class gender	Frail		count (%) Prefrail		Robust	
	Male	Female	Male	Female	Males	Female
Age groups						
(35, 50]	126 (1.7%)	162 (2.2%)	177 (2.4%)	223 (3.0%)	632 (8.5%)	348 (4.7%)
(50, 65]	160 (2.2%)	204 (2.8%)	256 (3.5%)	291 (3.9%)	842 (11.4%)	545 (7.4%)
(65, 80]	270 (3.6%)	357 (4.8%)	343 (4.6%)	433 (5.9%)	768 (10.4%)	745 (10.1 %)
(80, 95]	163 (2.2%)	133 (1.8%)	36 (0.5%)	54 (0.7%)	48 (0.6%)	71 (1%)
(95, 110]	11 (0.1%)	2 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Table 1: Age, gender and frailty status of participants.

index items using the data (data-driven approach).

This study aims to prioritize the questions in KCL according to their importance and contribution to the decision-making process. and subsequently, create a shorter version of KCL and a more straightforward interpretation using Explainable AI.

## 2. Methods

### 2.1. Subjects

The subjects of this study were randomly selected among people of 40 years or older in Settsu (February – March 2019) and Hannan (January – February 2020) cities from Osaka prefecture, Japan. The participants were stratified by Age and Sex, and they were evenly distributed geographically using school zones[30, 31]. The study questionnaire was submitted by mail. The total number of post-mails sent was 10000 in Settsu city and 8000 in Hannan city. The number of participants was 5809 (58 %) and 4801 (60 %), respectively. Later on, the responses with incomplete answers were excluded, and the total number went for the subsequent analysis steps was 7400 responses. The age and gender of the participants is summarized in table 1. Scores of less than 4 were considered robust, from 4 to 6 were considered pre-frail, and scores of 7 or more were considered frail.

The data analyzed in this paper was the KCL questionnaire and Food intake frequency survey.

## 2.2. Machine learning feature selection

To gain insight into how individual KCL questionnaire question informs the classification process of subjects, we followed a novel approach of combining machine learning models building with cross validation evaluation, then feature selection and models building again using those features in order. the code of this approach is available in <https://github.com/attayeb/proml>.

We built multiple machine learning models (Random forest, Gradient boost, SVM, Logistic regression, Decision tree, and Gaussian Naive Bayes) and validated them. First, shapely values were calculated for each model, and the top features were identified. Then, multiple models are created for each machine learning algorithm using the top two features at the beginning, with ten folds of cross-validations. Next, the average of Matthews correlation coefficients was calculated for all the cross-validation folds. Then progressively adding one feature at a time, repeat the same steps. Finally, the top-performing models were visualized to show how features contribute to the accuracy of the prediction models (Check Algorithm 1). The data used for building the models were all the data set, with all variables. Three models were built, to differentiate between all pairs of subject frailty classes. The default parameters according to sci-kit learn package were used. We chose Matthews correlation coefficient because it is recommended in imbalanced data sets [32].

**Input:** Questionnaire data; list of ML models

**Result:** Summarized Models performance

**begin**

**foreach** *model in machine learning models* **do**

        Train the model using all the questionnaire data;

        Identify selected features using Shapely values;

**for**  $i \leftarrow 2$  to 20 **do**

            Build a machine learning model using top  $i$  features;

            measure the performance of the model using Cross validation;

**end**

**end**

    Concatenate the results of the different model performances;

    Plot summarized figure;

**end**

**Algorithm 1:** Machine learning models evaluation

### 2.3. Subgroup detection

The original classification of subjects was determined based on the KCL questionnaire criteria, which utilizes the same data we used to build our classification models and feature selection process. This requires further evaluation if the data is possibly clusterable to clusters that corresponds at least partially with the criteria results. To do that, we examined if the questionnaire results could be clustered using unsupervised clustering algorithms. We reduced the data dimensions to 2 dimensions and plotted them. We tried PCA (Principal component analysis) [33], t-SNE (t-Distributed Stochastic Neighbor Embedding) [34] and UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) [35]. Before running these algorithms dimension reduction, we applied quantile transformation for data scaling using scikit-learn [36] QuantileTransformer function.



#### 2.4. *Used packages*

All model training, validation, and feature selection was done in local system using Ubuntu 20.04. All the analysis was done in Python programming language version 3.9; For PCA, RandomForest, SVM, and logistic regression, we used scikit-learn package [36], for feature importance, we use SHAP (SHapley Additive exPlanations) package [37].

### 3. Results

#### 3.1. *Progressive feature contribution assessment*

To understand the impact of the questions on the classification accuracy, we applied shapely values algorithm to prioritize these questions depending on their contribution to the decision of the prediction model, then we examined the performance of the model by adding the features one by one progressively starting with the top features with the greatest importance. This approach demonstrates the importance of any question in the final classification of the case. Addition of one question at a time make it possible to decide at what number of questions the accuracy of the prediction using machine learning is enough.

Classifying subject into either Robust or Frail does not require many questions to reach reasonable performance measured by Matthews' correlation coefficient (Figure 1). Only four questions were required: namely, "Do you visit home of friends?", "Are you able to go upstairs without using handrails or the wall for support?", "Do you feel anxious about falling when you walk?" and "In the past two weeks, have you felt exhausted in no apparent reasons?" can successfully differentiate Robust from frail classes. The accuracy using these 4 questions with Gradient boost model with 0.725 of Matthews correlation coefficient values.

Those four questions were the most commonly selected over Decision Tree, Gradient Boost, and logistic regression.

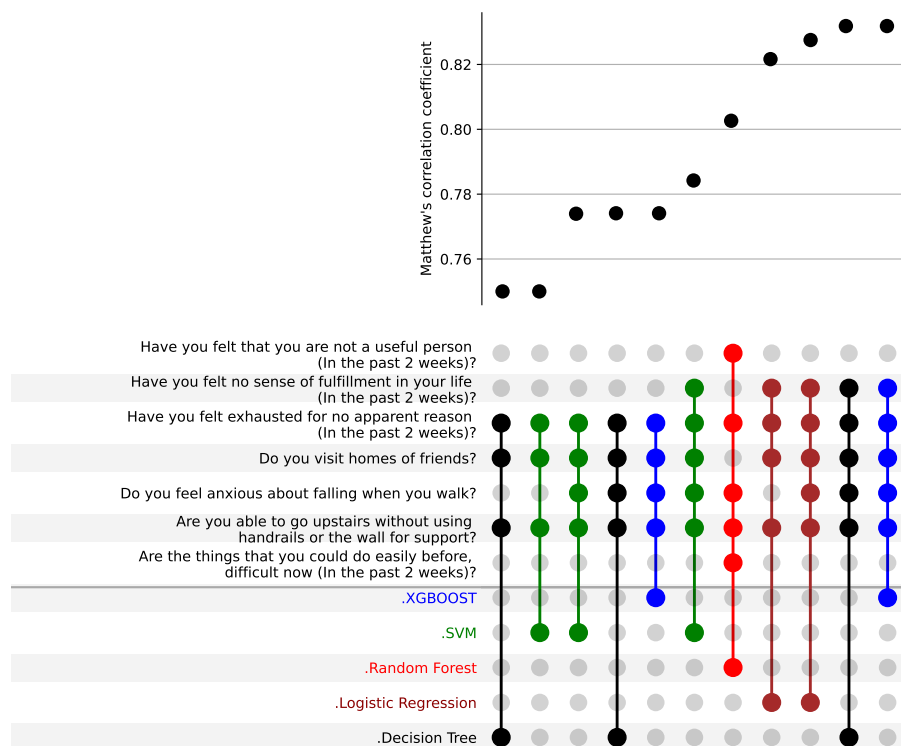


Figure 1: Progressive feature contribution assessment for Robust vs Frail. Small number of features are required to differentiate (as little as 4 questions can give prediction with 0.75 of Matthews correlation coefficient).

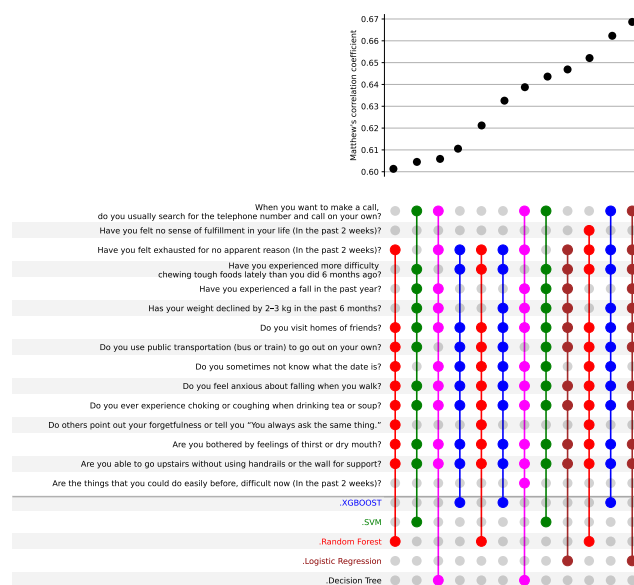


Figure 2: Progressive feature contribution assessment in differentiating Robust from Pre-frail classes, color represents the machine learning modalities and the questions was added gradually to find the best combination of questions that lead to better prediction. (MCC: Matthews Correlation Coefficient).

A higher number of questions is required to accurately identify subjects as Robust or Pre-frail (as shown in Figure 2). Around ten questions are required to have MCC values of more than 0.6; those questions were shared among machine learning algorithms demonstrating the consistency of their relevance to the decision process.

Frail with Pre-frail is also a challenging task. It can be differentiated by asking a higher number of questions, including “Do you use public transportation”, “Are you bothered by the feeling of thirst or dry mouth?” and “In the past two weeks, are the things that you could do easily before, difficult now?”. The accuracy was around more than 0.65 Matthews correlation coefficient.

### 3.2. Suggested short Questionnaire

Creating a shorter version of KCL questionnaire can help increase the usability of KCL questionnaire, having less number of question makes the screening less time consuming and more straight forward. To do that, we chose the top 4 crucial questions

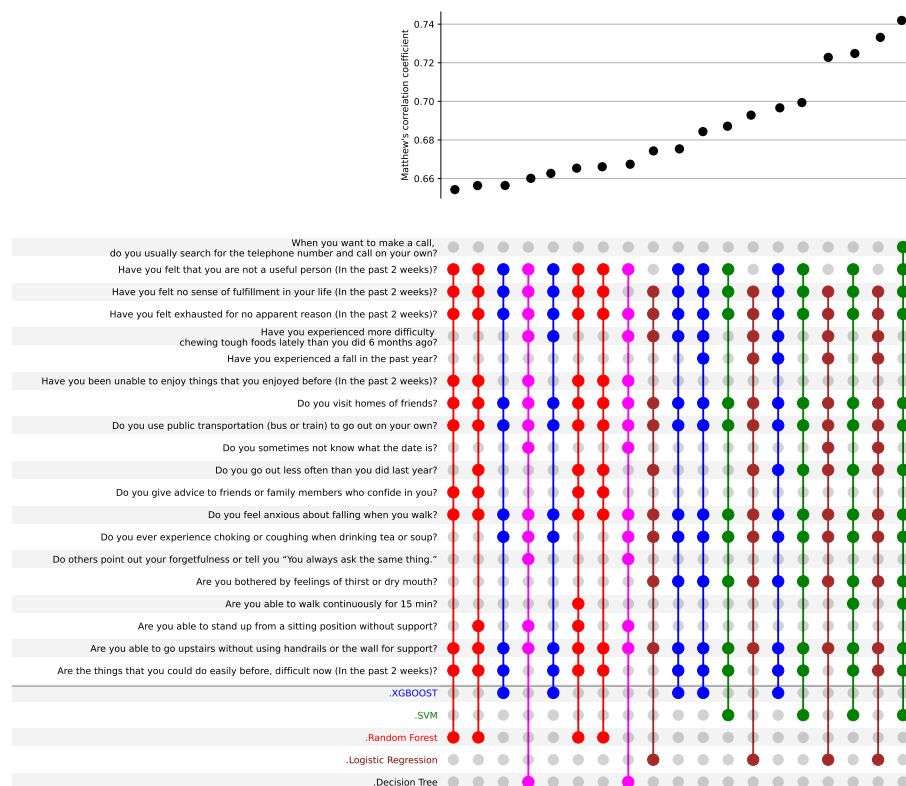


Figure 3: Progressive feature contribution assessment in differentiating frail from Pre-frail classes, color represents the machine learning modalities and the questions were added gradually to find the best combination of questions that lead to better prediction. (MCC: Matthews Correlation Coefficient).

identified in the Frail-Robust model, we built Random Forest model. We created a sham dataset with all possible combination of answers (16) and then calculated the probability of classification as robust (Check Table 2).

Do you visit homes of friends?	Are you able to go upstairs without using handrails or the wall for support?	Do you feel anxious about falling when you walk?	(In the past 2 weeks) Have you felt exhausted for no apparent reason?	Robustness Probability
✓	✓	✓	✓	72
✓	✓	✓	✗	99
✓	✓	✗	✓	99
✓	✓	✗	✗	100
✓	✗	✓	✓	16
✓	✗	✓	✗	86
✓	✗	✗	✓	54
✓	✗	✗	✗	100
✗	✓	✓	✓	12
✗	✓	✓	✗	79
✗	✓	✗	✓	50
✗	✓	✗	✗	97
✗	✗	✓	✓	0
✗	✗	✓	✗	30
✗	✗	✗	✓	8
✗	✗	✗	✗	75

Table 2: Short version of Kihon Questionnaire. (✓): answering Yes, (✗): answering No. Four questions are selected, the “Robustness probability” columns shows the probability of being robust using the answers of these 4 questions.

### 3.3. Subgroup discovery

To further reinforce our conclusion of the selected questions, we showed the data items collected can be clustered to groups corresponds to the KCL criteria classification (Figures 4, 5, 6). The segregation is more prominent when it comes to Frail compared to the pre-frail and Robust classes, however, the separation between Pre-frail and Robust was less obvious (Figures 4C, 5C, 6C).

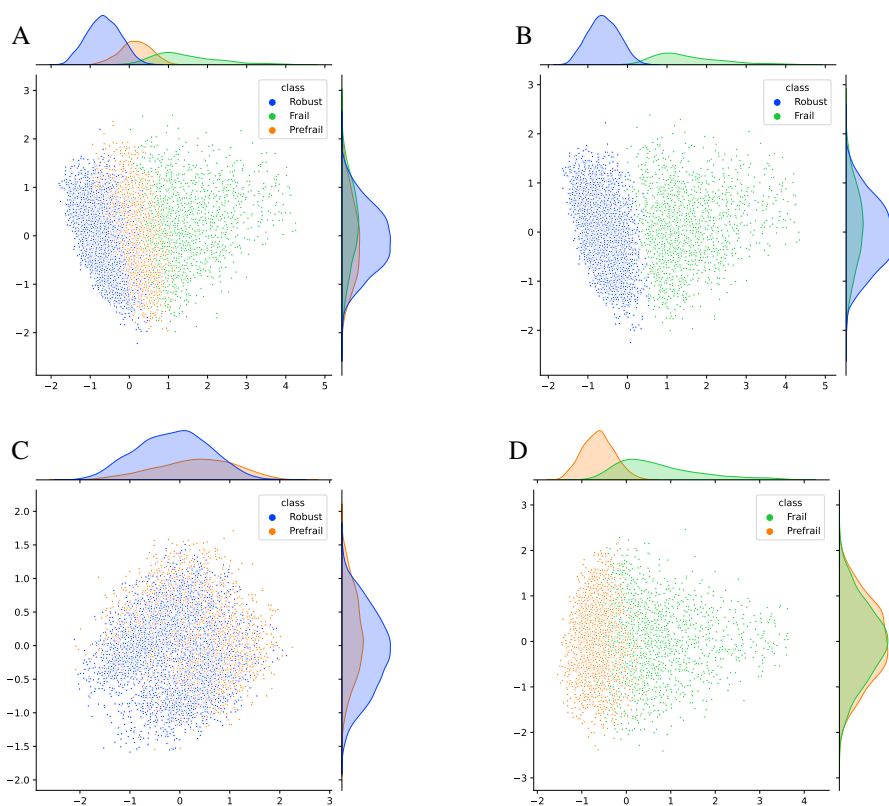
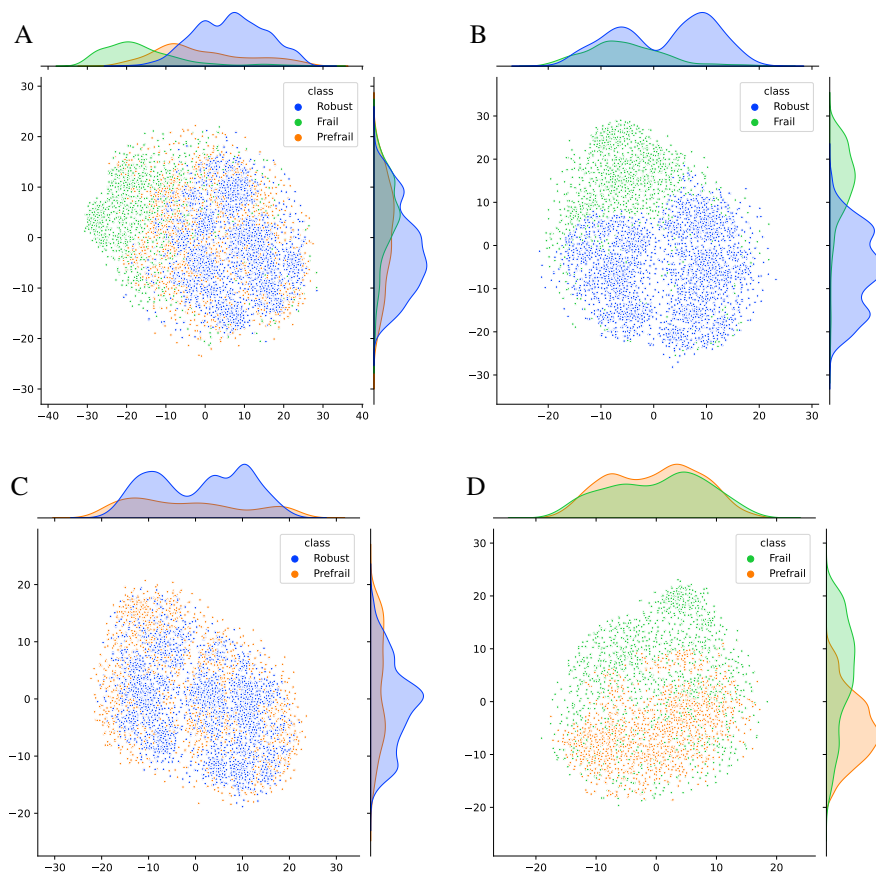
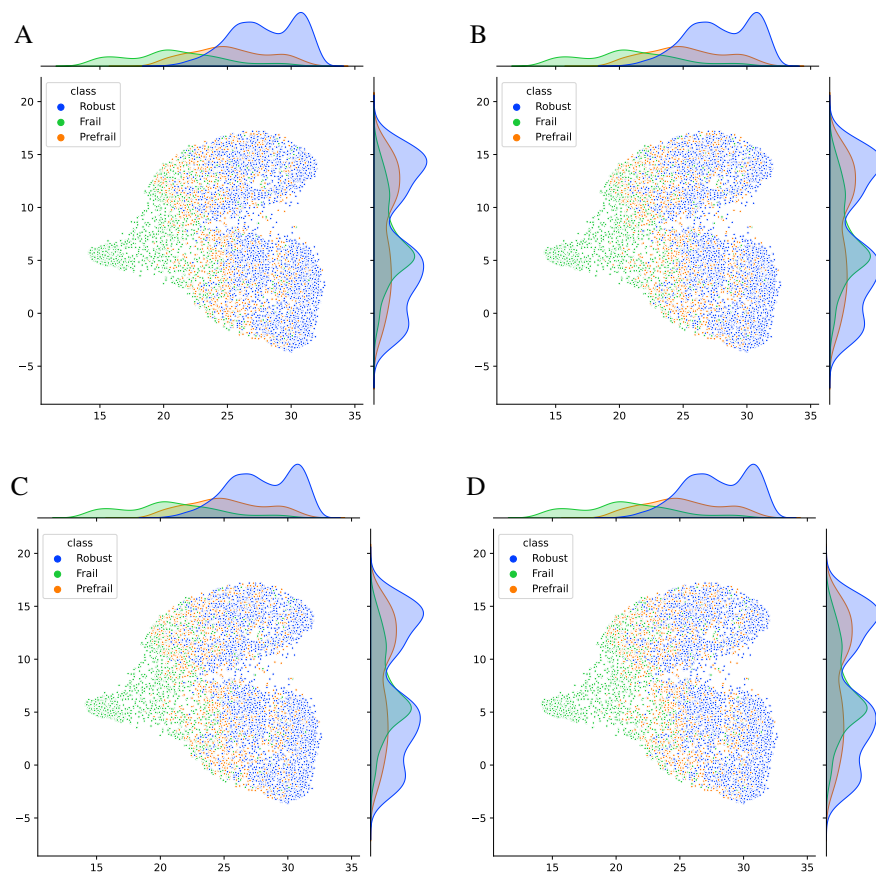


Figure 4: **Distribution of subjects in two dimensional PCA (principle component analysis) of questionnaire answers regarding to their class.** A. all three classes, B. Robust and Frail classes, C. Robust and Prefrail, D. Frail and prefrail. Principle component analysis was measured with 2 components. A, B, and D showed good subgroup discovery in contrast to C. (top and right side shows density plot of first and second components respectively)



**Figure 5: Distribution of subjects in two dimensional t-Distributed Stochastic Neighbor Embedding (t-SNE) of questionnaire answers regarding to their class. A. all three classes, B. Robust and Frail classes, C. Robust and Prefrail, D. Frail and prefrail. t-SNE embedding was plotted in 2 components. A, B, and D showed good subgroup discovery in contrast to C. (top and right side shows density plot of first and second dimensions respectively)**



**Figure 6: Distribution of subjects in two dimensional Uniform Manifold Approximation and Projection (UMAP) of questionnaire answers regarding to their class. A. all three classes, B. Robust and Frail classes, C. Robust and Prefrail, D. Frail and prefrail. UMAP was measured and plotted in 2 dimensions. A, B, and D showed good subgroup discovery in contrast to C. (top and right side shows density plot of first and second dimensions respectively)**



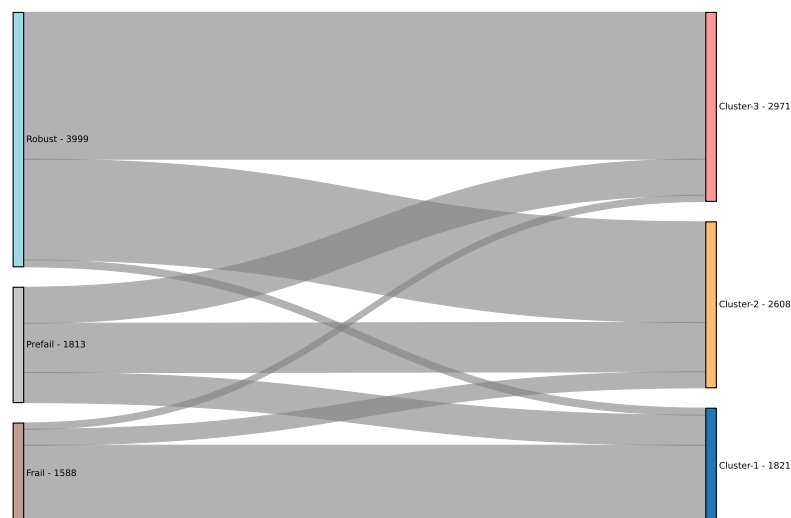


Figure 7: Kmeans clustering compatibility with the original frailty classes. Most of robust subjects belong to cluster 3, and Frail to cluster 1, however, prefrail was not clearly clustered in one cluster, it was dispersed between the three clusters.

### 3.4. Class-cluster overlap

To examine if the questionnaire can correctly classify the data using unsupervised clustering method and at the same time understand how robust is that, we clustered the data to three clusters using Kmeans clustering algorithm. The close examination of the resulting clusters (Figure 7) shows that majority of cluster 3 belongs to Robust class, and the majority of cluster 1 belongs to Frail class, however, the Pre-frail group is divided between the three clusters, and cluster2 contained mainly subjects from Robust and Pre-frail. This observation suggests that the questionnaire robustness was higher when it comes to differentiate Robust from frail, however determination of Pre-frail class is more difficult.

To provide a more comprehensive understanding of sample clustering and to compare the resulting clusters with the original classification, we re-analyzed the data using the Kmeans algorithm with 10 clusters. This approach allowed for a more flexible and informative depiction compared to the initial clustering of three clusters. Additionally,

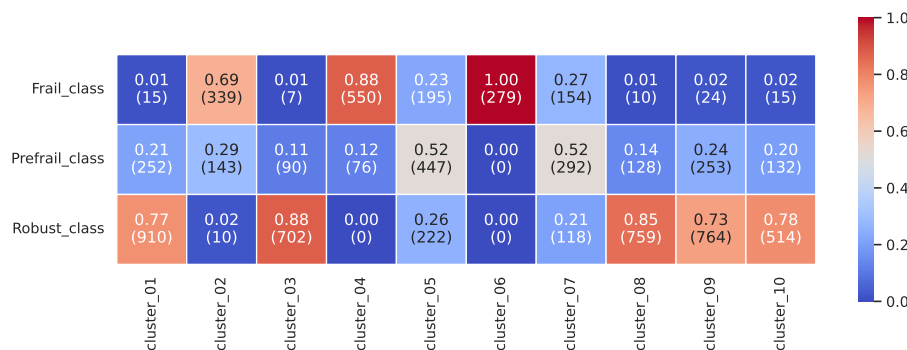


Figure 8: Overlap heatmap between three classes and 10 data clusters. Clusters [1, 3, 8, 9 and 10] have higher overlap with Robust class, meanwhile clusters [2, 4, 6] overlaps more with frail class. Half of clusters [5, 7] members belong to prefrail class, and the other half is equally distributed between the other two classes.

we calculated the overlap between the resulting clusters and the original classes (Figure 8). There are no substantially shared clusters between Robust and Frail, i.e. any given cluster should belong to either one of them but not both. but when it comes to pre-frail, There is no single cluster belongs to Pre-frail with 0.5 or more overlap. They are always shared between pre-frail and robust, or pre-frail and Frail.

In Figure 9, we clustered the data points to two clusters, plot them after reducing the dimensions using factor analysis. Most of Robust and Prefrail were clustered in one cluster, and frail was clustered alone. We calculated Matthews correlation coefficient after merging Prefrail and Robust, the clusters shows a value of 0.835, which suggested very high similarity between the original classes of data and the clusters.

These results suggested that the questions in the questionnaire was best in differentiating Frail from Robust, and at the same time, they could separate between Pre-frail and the other two classes with less accuracy.

#### 4. Discussion

In this paper, we ranked the KCL questions by their significance in determining frailty status. We also expanded these results by proposing a simplified KCL questionnaire that can be used for screening purposes through a novel approach that integrates

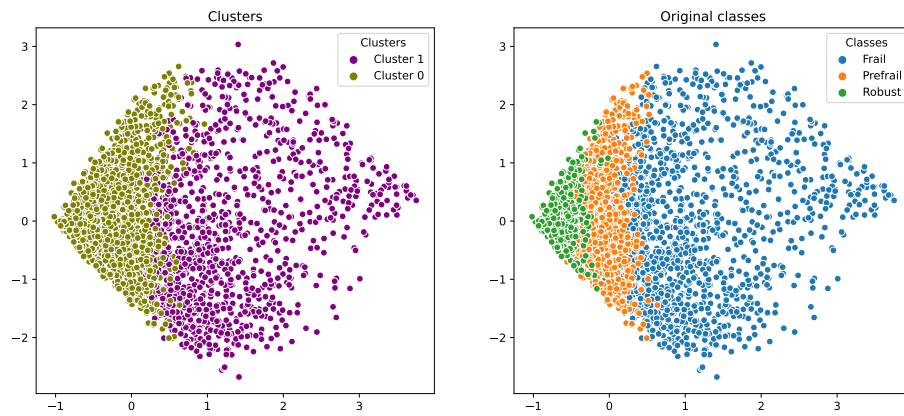


Figure 9: Clustering the data to two clusters using KMeans, and presenting them in two dimensions after Factor analysis dimension reduction. Left side shows the clusters, and the right side shows the original classes identified by KCL criteria.

feature selection and the gradual construction and evaluation of machine learning models. We also validated the usability of the data collected using the KCL questionnaire by unsupervised clustering that produced two distinctive clusters (Frail and Robust) and the third is an intermediate cluster (Pre-frail). The ability of KCL to produce these three clusters suggests the robustness of KCL as an indicator of frailty.

The responders of this questionnaire were assigned to three groups of robust, frail and pre-frail following simple criteria that considers the number of questions answered. To validate if this survey is sensitive, we cluster the subjects' responses using multiple unsupervised methods. The objective of this clustering is to confirm if the data can be adequately classified without any intervention. Figures (4, 5, 6, 7, and 8) demonstrate the clear separation between the clusters of frail and robust subjects, with the pre-frail class occupying the intermediate position. That is more prominent in figure 8, where no cluster has a similar number of individuals shared among the three groups or particularly among frail and robust. This observation implies that the questionnaire was designed and interpreted in a manner that produces reliable results. Had some questions were not carefully chosen or not being informative, the resulting clusters would contain a cluster or more with majority of both robust and frail.

The early diagnosis of frailty paves the road to recovery and the patient’s protection against quick deterioration. Therefore, creating a frailty test that can be easily used for screening is crucial [8].

Frailty research has extensively utilized machine learning [25–29] and explainable AI [38] That can shed light on the mechanisms of diseases and identify biomarkers that can be used to diagnose or find a treatment for a specific medical condition. Moreover, it can be extended to other applications, such as understanding the prognostic factors or severity modulators[39].

Using machine learning and explainable AI allowed us to produce a shorter version of the KCL questionnaire with reasonable accuracy. Moreover, The questions selected (Table: 2) reflect both physical and psychological status. For instance, the question “(4) Do you visit homes of friends?” reflects the social well-being of the subject, which includes psychological status but it also extends to motor activities. Conversely, the subject’s inability or disinterest in visiting friends’ homes could stem from either or both physical and psychological limitations. The second question about going upstairs without using handrails or the wall for support is mainly a physical indicator. However, if the subject has safety concerns or doubts about their abilities, they may still use handrails, thus also serving as a psychological indicator. This relation between physical and mental health is further evident in the question “(10) Do you feel anxious about falling when you walk?” which similarly reflects both aspects of the subject’s health.

The data showed in this paper suggest that using a combination of comprehensive questions showed a better indication of frailty than purely physical or psychological questions such as “Are you able to stand up from a sitting position without support?” or “Are you able to walk continuously for 15 min?”.

The accuracy of the result simplified questionnaire is lower than the accuracy if we use the entire questionnaire. Also, our method showed it is difficult to predict the pre-frail state using a smaller number of KCL questions. This limitation arises from

the fact of resemblance between pre-frail and Robust.

## 5. Conclusions

In this paper we examined the performance of KCL questionnaire using a novel machine learning-feature selection approach. We identified and prioritized the questions according to their contribution in the classification decision made. Moreover, we suggest a shortened version of KCL questionnaire with 4 questions: (4) Do you visit homes of friends?, (6) Are you able to go upstairs without using handrails or the wall for support? (10) Do you feel anxious about falling when you walk?, and (25) (In the past two weeks) Have you felt exhausted for no apparent reason?. This short version of KCL can be used in clinical settings for the prediction of frailty status with reasonable accuracy.

## Abbreviations

KCL: Kihon CheckList;

## Acknowledgment

### Author contribution:

**Conceptualization:** Motohiko Miyachi, Michihiro Araki, **Data Collection:** Daiki Watanabe, Nobuo Nishi, Takashi Nakagata, Tsukasa Yoshida, Motohiko Miyachi, **Data Analysis:** Masaki Yamamoto, Attayeb Mohsen, **Writing – Original draft:** Attayeb Mohsen, **Writing – Review and editing:** Attayeb Mohsen, Tsukasa Yoshida, Agustin Martin Morales, Michihiro Araki, Motohiko Miyachi, All authors checked the final draft and approved the submission.

## Funding

This research is supported by: Japan Science and Technology Agency: COI-NEXT (grant number JPMJPF2018 to M.A.)

## References

- [1] R. Romero-Ortuno and D. O'Shea. "Fitness and frailty: opposite ends of a challenging continuum! Will the end of age discrimination make frailty assessments an imperative?" en. In: *Age Ageing* 42.3 (May 2013), pp. 279–280.
- [2] E. Dent, C. Lien, W. S. Lim, et al. "The Asia-Pacific Clinical Practice Guidelines for the management of frailty". en. In: *J. Am. Med. Dir. Assoc.* 18.7 (July 2017), pp. 564–575.
- [3] E. O. Hoogendijk, J. Afilalo, K. E. Ensrud, et al. "Frailty: implications for clinical practice and public health". en. In: *Lancet* 394.10206 (Oct. 2019), pp. 1365–1375.
- [4] L. P. Fried, C. M. Tangen, J. Walston, et al. "Frailty in Older Adults: Evidence for a Phenotype". In: *The Journals of Gerontology: Series A* 56.3 (Mar. 2001), pp. M146–M157. ISSN: 1079-5006. DOI: 10.1093/gerona/56.3.M146. eprint: <https://academic.oup.com/biomedgerontology/article-pdf/56/3/M146/9743160/M146.pdf>.
- [5] T. Lancet. "Bringing frailty into all realms of medicine". In: *The Lancet* 394.10206 (Oct. 2019), p. 1298. DOI: 10.1016/s0140-6736(19)32279-2.
- [6] J. E. Morley, B. Vellas, G. A. van Kan, et al. "Frailty Consensus: A Call to Action". In: *Journal of the American Medical Directors Association* 14.6 (June 2013), pp. 392–397. DOI: 10.1016/j.jamda.2013.03.022.

- [7] M. Rosenberg, K. Kondo, N. Kondo, H. Shimada, and H. Arai. “Primary care approach to frailty: Japan’s latest trial in responding to the emerging needs of an ageing population”. In: *Integrated Healthcare Journal* 2.1 (Oct. 2020), e000049. DOI: 10.1136/ihj-2020-000049.
- [8] A. Lacas and K. Rockwood. “Frailty in primary care: a review of its conceptualization and implications for practice”. In: *BMC Medicine* 10.1 (Jan. 2012). DOI: 10.1186/1741-7015-10-4.
- [9] M. Ohashi, T. Yoda, N. Imai, et al. “Five-year longitudinal study of frailty prevalence and course assessed using the Kihon Checklist among community-dwelling older adults in Japan”. en. In: *Scientific Reports* 11.11 (June 2021), p. 12399. ISSN: 2045-2322. DOI: 10.1038/s41598-021-91979-6.
- [10] K. Rockwood, K. Stadnyk, C. MacKnight, et al. “A brief clinical instrument to classify frailty in elderly people”. en. In: *Lancet* 353.9148 (Jan. 1999), pp. 205–206.
- [11] S. D. Searle, A. Mitnitski, E. A. Gahbauer, T. M. Gill, and K. Rockwood. “A standard procedure for creating a frailty index”. In: *BMC Geriatrics* 8.1 (Sept. 2008), p. 24. ISSN: 1471-2318. DOI: 10.1186/1471-2318-8-24.
- [12] H. Arai and S. Satake. “English translation of the Kihon Checklist”. In: *Geriatrics & Gerontology International* 15.4 (Apr. 2015), pp. 518–519. DOI: 10.1111/ggi.12397.
- [13] P. Y. S. Sampaio, R. A. C. Sampaio, M. Yamada, and H. Arai. “Systematic review of the Kihon Checklist: Is it a reliable assessment of frailty?” In: *Geriatrics & Gerontology International* 16.8 (July 2016), pp. 893–902. DOI: 10.1111/ggi.12833.
- [14] T. Suzuki. *The manual on evaluation of life function for preventing long-term care*. Article in Japanese. 2021.

- [15] Japanese Ministry of Health, Labour and Welfare. *Kihon Checklist no kangae-kata ni tsuite (in Japanese). Concept of Kihon Check list. (2006)*. Article in Japanese. 2006. URL: <https://www.mhlw.go.jp/topics/2007/03/dl/tp0313-1a-11.pdf>.
- [16] S. Satake, K. Senda, Y.-J. Hong, et al. “Validity of the Kihon Checklist for assessing frailty status”. en. In: *Geriatr. Gerontol. Int.* 16.6 (June 2016), pp. 709–715.
- [17] R. C. Ambagtsheer, R. Visvanathan, E. Dent, et al. “Commonly used screening instruments to identify frailty among community-dwelling older people in a general practice (primary care) setting: A study of diagnostic test accuracy”. en. In: *J. Gerontol. A Biol. Sci. Med. Sci.* 75.6 (May 2020), pp. 1134–1142.
- [18] D. Watanabe, T. Yoshida, Y. Watanabe, et al. “Validation of the Kihon Checklist and the frailty screening index for frailty defined by the phenotype model in older Japanese adults”. en. In: *BMC Geriatr.* 22.1 (June 2022), p. 478.
- [19] D. Watanabe, T. Yoshida, Y. Yamada, et al. “Combined use of two frailty tools in predicting mortality in older adults”. en. In: *Sci. Rep.* 12.1 (Sept. 2022), p. 15042.
- [20] A. Tao, Y. Zhang, X. Qiu, H. Arai, and Q. Wang. “Simplified Chinese translation of the Kihon checklist”. en. In: *Geriatr. Gerontol. Int.* 20.6 (June 2020), pp. 643–644.
- [21] I.-Y. Jang, H.-W. Jung, C. K. Lee, et al. “Korean version of the Fatigue, Resistance, Ambulation, Illnesses and Loss of weight questionnaire versus the Modified Kihon Checklist for Frailty Screening in Community-Dwelling Older Adults: The Aging Study of PyeongChang Rural Area”. en. In: *Geriatr. Gerontol. Int.* 17.11 (Nov. 2017), pp. 2046–2052.



- [22] M. E. Esenkaya, O. Dokuzlar, P. Soysal, et al. “Validity of the Kihon Checklist for evaluating frailty status in Turkish older adults”. en. In: *Geriatr. Gerontol. Int.* 19.7 (July 2019), pp. 616–621.
- [23] P. Y. Sewo Sampaio, R. A. C. Sampaio, M. Yamada, M. Ogita, and H. Arai. “Validation and translation of the Kihon Checklist (frailty index) into Brazilian Portuguese”. en. In: *Geriatr. Gerontol. Int.* 14.3 (July 2014), pp. 561–569.
- [24] A. Maseda, L. Lorenzo-López, R. López-López, H. Arai, and J. C. Millán-Calenti. “Spanish translation of the Kihon Checklist (frailty index)”. en. In: *Geriatr. Gerontol. Int.* 17.3 (Mar. 2017), pp. 515–517.
- [25] J. A. V. d. Sousa, M. H. Lenardt, C. R. B. Grden, et al. “Physical frailty prediction model for the oldest old”. In: *Revista Latino-Americana de Enfermagem* 26.0 (Sept. 2018). ISSN: 1518-8345. DOI: 10.1590/1518-8345.2346.3023. URL: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0104-11692018000100346&lng=en&tlng=en](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-11692018000100346&lng=en&tlng=en).
- [26] A. P. Hassler, E. Menasalvas, F. J. García-García, L. Rodríguez-Mañas, and A. Holzinger. “Importance of medical data preprocessing in predictive modeling and risk factor discovery for the frailty syndrome”. In: *BMC Medical Informatics and Decision Making* 19.1 (Feb. 2019). DOI: 10.1186/s12911-019-0747-6.
- [27] S. Aponte-Hao, S. T. Wong, M. Thandi, et al. “Machine learning for identification of frailty in Canadian primary care practices”. In: *International Journal of Population Data Science* 6.1 (Sept. 2021). DOI: 10.23889/ijpds.v6i1.1650.
- [28] K.-M. Kuo, P. C. Talley, M. Kuzuya, and C.-H. Huang. “Development of a clinical support system for identifying social frailty”. In: *International Journal of Medical Informatics* 132 (Dec. 2019), p. 103979. DOI: 10.1016/j.ijmedinf.2019.103979.

- [29] S.-H. Tseng, L.-K. Liu, L.-N. Peng, et al. “Development and Validation of a Tool to Screen for Cognitive Frailty Among Community-Dwelling Elders”. In: *The journal of nutrition, health & aging* 23.9 (Aug. 2019), pp. 904–909. DOI: 10.1007/s12603-019-1235-5.
- [30] T. Yoshida, D. Watanabe, T. Nakagata, et al. “[Prevalence of frailty and its related factors in community-dwelling middle-aged and elderly adults in Settsu and Hannan cities in Osaka prefecture]”. Article in Japanese. In: *[Nihon Koshu Eisei Zasshi] Japanese Journal of Public Health* 68.8 (Aug. 2021), pp. 525–537. ISSN: 0546-1766. DOI: 10.11236/jph.20-111.
- [31] D. Watanabe, T. Yoshida, T. Nakagata, et al. “Factors associated with sarcopenia screened by finger-circle test among middle-aged and older adults: a population-based multisite cross-sectional survey in Japan”. In: *BMC Public Health* 21.1 (Apr. 2021), p. 798. ISSN: 1471-2458. DOI: 10.1186/s12889-021-10844-3.
- [32] D. Chicco and G. Jurman. “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation”. en. In: *BMC Genomics* 21.1 (Dec. 2020), p. 6. ISSN: 1471-2164. DOI: 10.1186/s12864-019-6413-7.
- [33] I. T. Jolliffe and J. Cadima. “Principal component analysis: a review and recent developments”. en. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065 (Apr. 2016), p. 20150202. ISSN: 1364-503X, 1471-2962. DOI: 10.1098/rsta.2015.0202.
- [34] L. van der Maaten and G. Hinton. “Visualizing Data using t-SNE”. In: *Journal of Machine Learning Research* 9.86 (2008), pp. 2579–2605. URL: <http://jmlr.org/papers/v9/vandermaaten08a.html>.
- [35] L. McInnes, J. Healy, and J. Melville. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. 2018. DOI: 10.48550/ARXIV.1802.03426. arXiv: 1802.03426.

- [36] F. Pedregosa, G. Varoquaux, A. Gramfort, et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [37] S. M. Lundberg and S.-I. Lee. “A Unified Approach to Interpreting Model Predictions”. In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, et al. Curran Associates, Inc., 2017, pp. 4765–4774. URL: <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>.
- [38] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis. “Explainable AI: A Review of Machine Learning Interpretability Methods”. In: *Entropy* 23.1 (Dec. 2020), p. 18. DOI: 10.3390/e23010018.
- [39] D. Jaber, H. Hajj, F. Maalouf, and W. El-Hajj. “Medically-oriented design for explainable AI for stress prediction from physiological measurements”. In: *BMC Medical Informatics and Decision Making* 22.1 (Feb. 2022). DOI: 10.1186/s12911-022-01772-2.